

Join the shortest queue among k parallel queues: tail asymptotics of its stationary distribution

Masahiro Kobayashi^a, Yutaka Sakuma^b and Masakiyo Miyazawa^a

^aDepartment of Information Sciences, Tokyo University of Science

^bHiroshima National College of Maritime Technology

October 21, 2012

Abstract

We are concerned with an M/M -type join the shortest queue (M/M -JSQ for short) with k parallel queues for an arbitrary positive integer k , where the servers may be heterogeneous. We are interested in the tail asymptotic of the stationary distribution of this queueing model, provided the system is stable. We prove that this asymptotic for the minimum queue length is exactly geometric, and its decay rate is the k -th power of the traffic intensity of the corresponding k server queues with a single waiting line. For this, we use two formulations, a quasi-birth-and-death (QBD for short) process and a reflecting random walk on the boundary of the $k + 1$ -dimensional orthant. The QBD process is typically used in the literature for studying the JSQ with 2 parallel queues, but the random walk also plays a key roll in our arguments, which enables us to use the existing results on tail asymptotics for the QBD process.

Keywords: Join the shortest queue, heterogeneous servers, stationary distribution, exactly geometric asymptotics, quasi-birth-and-death process, reflecting random walk.

1 Introduction

We consider a parallel queueing model in which customers join the shortest queue. If there are more than one queues whose lengths are shortest, then we assume tie break with equal probabilities. We denote the number of queues by k (≥ 2). It is assumed that customers arrive according to a Poisson process, and each queue has a single server, and its service times are *i.i.d* with an exponential distribution. Here, those servers may have different mean service times, that is, they may be heterogeneous. We refer to this queueing model as an M/M -type join the shortest queue (M/M -JSQ for short).

We are interested in the stationary distribution of the queue length for the M/M -JSQ. However, its analytic derivation is known to be hard, and theoretical interests have been

directed to the tail asymptotic of the stationary distribution. For $k = 2$, this problem has been well studied. Kingman studied the M/M -JSQ with 2 parallel queues having homogeneous servers. He proved that the stationary distribution of the minimum queue length has an exactly geometric asymptotics and its decay rate is equal to the square of traffic intensity of the corresponding $M/M/2$ queue with a single waiting line, where exactly geometric asymptotics means that the tail probability is asymptotically proportional to a geometric function. Many researchers obtained similar geometric asymptotics for more general models with two parallel queues. The case of heterogeneous servers is studied in [15]. Foley and McDonald [2] considered the generalized shortest queue which has a Poisson stream dedicated to each queue in addition to a Poisson stream which chooses the shortest queue. They obtained the stability condition and exactly geometric asymptotic under some extra conditions. This model was further studied in [6, 8]. In particular, Miyazawa [8] described the generalized shortest queue by a two sided double quasi-birth-and-death (QBD) process, and derived the tail decay rates without any extra condition. Sakuma [14] considered two parallel queues with a common phase type service time distribution and a Markov modulated arrivals, and derived exactly geometric asymptotic under a certain condition.

All those studies for JSQ assume two parallel queues. We consider the case where there are more than two parallel queues, that is, k queues with $k \geq 3$. Puhalskii and Vladimirov [12] studied the tail asymptotics for much more general k parallel queues in which there are multiple classes of customers who can only choose the shortest queue among queues assigned to them. They derived the large deviations principle for this generalized model. However, they do not answer any explicit asymptotics even for $k = 2$. Sakuma [13] also studied k parallel queues with JSQ discipline, but this model is allowed to have jockeying when the maximum difference among queues is greater than a given threshold level. Due to this jockeying assumption, the problem is reduced to one dimensional queue, and the standard technique can be applied to get tail asymptotics.

For the M/M -JSQ with k parallel queues, it is easy to guess that the tail decay rate of the stationary distribution for the minimum queue length is the k -th power of the traffic intensity of the corresponding k server queue with a single waiting line because all the queues should be balanced. However, there has been no proof for $k \geq 3$ as far as we know. This may be because there is no satisfactory method for tail asymptotics on the stationary distribution for more than two correlated queues.

The aim of this paper is to prove this conjecture. This will be done by obtaining exactly geometric asymptotic. For this, we employ two formulations. We first consider the exact geometric asymptotics using a discrete time QBD process, which has two components, level and background. The level is one dimensional, and represents a characteristics of interest for tail asymptotics. The background state has all the information for the process to be Markov. We define the level by the minimum queue length for the M/M -JSQ with k parallel queues while background state is a set of differences between the queues lengths and their minimum. Then, we have the QBD process, and the stationary distribution is represented by the so called matrix geometric solution. This solution can be used to derive the exactly geometric asymptotic. However, we need to verify certain conditions for this derivation which are not always easy to verify. In particular, one of them is involved with the tail asymptotic of the marginal distribution at level zero, which is generally unknown.

To overcome this difficulty, we use another formulation. We take the same state space as that of the QBD process. Thus, each state is a $k + 1$ dimensional vector at least one of whose entries vanishes. We consider this process as a reflecting random walk on the boundary of the $k + 1$ dimensional nonnegative integer orthant. This boundary is composed of $2(2^k - 1)$ faces depending on which entries vanish. This random walk provides us a different tool for solving the tail asymptotic problem. For this, we use moment generating functions for describing the stationary equation instead of the matrix geometric solution. Of course, it is very hard to analytically derive the moment generating function of the multidimensional stationary distribution. Instead of doing so, we only consider its convergence domain, similarly to our recent papers [4, 9]. We can not find the whole domain, but can get sufficient information to apply the tail asymptotic result of the QBD process.

The rest of this paper is organized as follows. In Section 2, we formally introduce a Markov chain for the shortest queue and formulate it in two ways, the QBD process and reflecting random walk. We then present a main result, the exact geometric asymptotic for the M/M -JSQ with k parallel queues (Theorem 2.1). As a corollary of this result, we also derive a rough asymptotic for the marginal distribution of the minimum queue length (Corollary 2.1). In Section 3, we prove the main result using one proposition and five lemmas. The first two lemmas are on the QBD process, and proved in the appendix. The last lemma plays a key roll for our proof. It is proved in Section 4, using further lemmas. We give some concluding remarks in Section 5.

2 Modeling and exactly geometric asymptotics

We consider a queueing model with k parallel single server queues, where each waiting line has infinite capacity, and we index those k queues as $1, 2, \dots, k$. Customers arrive according to a Poisson process with rate λ and join the shortest queue, where ties are broken with equal probabilities when there are more than one shortest queues. At the i -th queue, the customers are served according to first-come first-served discipline, and their service times are independent and exponentially distributed with mean μ_i^{-1} . Thus, the servers may not be homogeneous. This queueing model is referred to as an M/M -type join the shortest queue (M/M -JSQ) with k parallel queues (see also Figure 1).

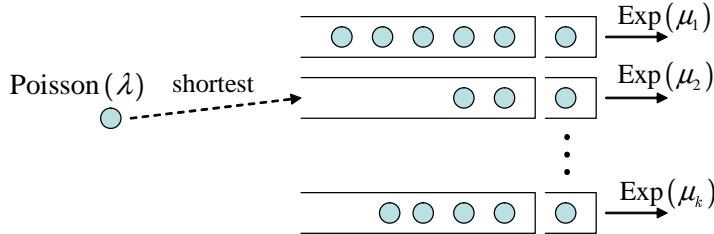


Figure 1: M/M -JSQ with k parallel queues

We denote the index set of queues by J , that is,

$$J = \{1, 2, \dots, k\}.$$

For each t and $i \in J$, let $L_i(t)$ be the number of customers in queue i including a customer being served, and let

$$\mathbf{L}(t) = (L_1(t), L_2(t), \dots, L_k(t)).$$

It is easy to see that $\{\mathbf{L}(t); t \in \mathbb{R}_+\}$ is a continuous time Markov chain with state space \mathbb{Z}_+^k , where \mathbb{R}_+ and \mathbb{Z}_+ are the sets of all nonnegative real numbers and integers, respectively. We denote the traffic intensity of this queueing model by

$$\rho = \frac{\lambda}{\sum_{i=1}^k \mu_i},$$

and assume that

$$\rho < 1, \tag{2.1}$$

which is known to be the stability condition (see, e.g., [2]). Since the total transition rate from each state is bounded by $\lambda + \sum_{i=1}^k \mu_i$, we can construct a discrete time Markov chain which has the same stationary distribution as that of $\{\mathbf{L}(t)\}$ by uniformization. We normalize $\lambda + \sum_{i=1}^k \mu_i$ without loss of generality as

$$\lambda + \sum_{i=1}^k \mu_i = 1. \tag{2.2}$$

We denote this discrete time Markov chain by $\{\mathbf{L}_\ell; \ell = 0, 1, \dots\}$, where

$$\mathbf{L}_\ell = (L_{\ell 1}, L_{\ell 2}, \dots, L_{\ell k}). \tag{2.3}$$

In this paper, we refer to this process as an original queue length process.

In the rest of this paper, we consider this discrete time process. The state transitions of \mathbf{L}_ℓ are a bit complicated because they depend on how $L_{\ell i}$'s are ordered. Thus, we describe it in a slightly different way. Let

$$M_\ell = \min_{i \in J} L_{\ell i}, \quad \mathbf{Y}_\ell = (Y_{\ell 1}, Y_{\ell 2}, \dots, Y_{\ell k}), \quad \ell = 0, 1, \dots,$$

where $Y_{\ell i} = L_{\ell i} - M_\ell$ for $i \in J$. Let $\{\mathbf{Z}_\ell; \ell = 0, 1, \dots\}$ be a pair of M_ℓ and \mathbf{Y}_ℓ , that is,

$$\mathbf{Z}_\ell = (M_\ell, \mathbf{Y}_\ell),$$

which is just another expression for $\{\mathbf{L}_\ell\}$. Obviously, $\{\mathbf{Z}_\ell; \ell = 0, 1, 2, \dots\}$ is a discrete time Markov chain which takes values in \mathbb{Z}_+^{k+1} .

The process $\{\mathbf{Z}_\ell\}$ is convenient because of two reasons. First, it can be considered as a quasi-birth-and-death process, QBD for short, if we view M_ℓ as level and \mathbf{Y}_ℓ as background state. Secondly, it can be considered as a reflecting random walk. This

simplifies our arguments while keeping accurate mathematical expressions in the boundary faces of $k + 1$ -dimensional nonnegative integer orthant.

Let us describe the process $\{\mathbf{Z}_\ell\}$ as a $k + 1$ -dimensional reflecting random walk. We first partition its state space. Let $\mathcal{K} \equiv \{U \subset J; U \neq \emptyset\}$. For $U \in \mathcal{K}$, we define the following boundary faces.

$$\begin{aligned}\mathcal{S}_{+U} &= \{\mathbf{u} = (u_0, u_1, u_2, \dots, u_k) \in \mathbb{Z}_+^{k+1}; u_0 \geq 1, u_i \geq 1, i \in J \setminus U, u_j = 0, j \in U\}, \\ \mathcal{S}_{0U} &= \{\mathbf{u} = (u_0, u_1, u_2, \dots, u_k) \in \mathbb{Z}_+^{k+1}; u_0 = 0, u_i \geq 1, i \in J \setminus U, u_j = 0, j \in U\}.\end{aligned}$$

Denote the boundary of \mathbb{Z}_+^{k+1} by $\partial\mathbb{Z}_+^{k+1}$, that is,

$$\partial\mathbb{Z}_+^{k+1} = \{\mathbf{u} = (u_0, u_1, \dots, u_k) \in \mathbb{Z}_+^{k+1}; \exists i \in \{0\} \cup J, u_i = 0\}.$$

Then, $\partial\mathbb{Z}_+^{k+1} = \cup_{U \in \mathcal{K}} (\mathcal{S}_{+U} \cup \mathcal{S}_{0U})$, and $\{\mathbf{Z}_\ell\}$ only stays on the boundary $\partial\mathbb{Z}_+^{k+1}$. Thus, the state space of \mathbf{Z}_ℓ is given by

$$\mathcal{S} = \cup_{U \in \mathcal{K}} (\mathcal{S}_{+U} \cup \mathcal{S}_{0U}).$$

Note that $\mathbf{Z}_\ell \in \mathcal{S}_{+U}$ if U is the set of indices of the shortest queues and $M_\ell \geq 1$. Similarly, $\mathbf{Z}_\ell \in \mathcal{S}_{0U}$ implies that U is the set of indices of the shortest queues and $M_\ell = 0$.

Let us consider the distribution of increment $\mathbf{Z}_{\ell+1} - \mathbf{Z}_\ell$. It only depends on the boundary face to which \mathbf{Z}_ℓ belongs. To give its distribution, we use the following notations. For $i \in J$, let \mathbf{e}_i be the k -dimensional row vector whose i -th entry is unit and the other entries vanish (e.g., $\mathbf{e}_1 = (1, 0, \dots, 0)$), and let $\mathbf{1}$ be the k -dimensional row vector whose all entries are units, i.e., $\mathbf{1} = (1, 1, \dots, 1)$. We denote the number of elements of set U by $|U|$. For each $U \in \mathcal{K}$, we define the random vector $\mathbf{X}^{(+U)}$ taking value in $(j, \mathbf{v}) \in \{0, -1, 1\}^{k+1}$ as follows, for $|U| = 1$,

$$\mathbb{P}(\mathbf{X}^{(+U)} = (j, \mathbf{v})) = \begin{cases} \lambda, & (j, \mathbf{v}) = (1, -\mathbf{1} + \mathbf{e}_i), i \in U, \\ \mu_i, & (j, \mathbf{v}) = (0, -\mathbf{e}_i), i \in J \setminus U \\ & \text{or } (j, \mathbf{v}) = (-1, \mathbf{1} - \mathbf{e}_i), i \in U, \\ 0, & \text{otherwise,} \end{cases} \quad (2.4)$$

and for $|U| \geq 2$,

$$\mathbb{P}(\mathbf{X}^{(+U)} = (j, \mathbf{v})) = \begin{cases} \frac{1}{|U|}\lambda, & (j, \mathbf{v}) = (0, \mathbf{e}_i), i \in U, \\ \mu_i, & (j, \mathbf{v}) = (0, -\mathbf{e}_i), i \in J \setminus U \\ & \text{or } (j, \mathbf{v}) = (-1, \mathbf{1} - \mathbf{e}_i), i \in U, \\ 0, & \text{otherwise.} \end{cases} \quad (2.5)$$

Similarly, let $\mathbf{X}^{(0U)}$ be the random vector such that, for $|U| = 1$

$$\mathbb{P}(\mathbf{X}^{(0U)} = (j, \mathbf{v})) = \begin{cases} \lambda, & (j, \mathbf{v}) = (1, -\mathbf{1} + \mathbf{e}_i), i \in U, \\ \mu_i, & (j, \mathbf{v}) = (0, -\mathbf{e}_i), i \in J \setminus U \\ & \text{or } (j, \mathbf{v}) = (0, \mathbf{0}), i \in U, \\ 0, & \text{otherwise,} \end{cases} \quad (2.6)$$

and for $|U| \geq 2$,

$$\mathbb{P}(\mathbf{X}^{(0U)} = (j, \mathbf{v})) = \begin{cases} \frac{1}{|U|}\lambda, & (j, \mathbf{v}) = (0, \mathbf{e}_i), i \in U, \\ \mu_i, & (j, \mathbf{v}) = (0, -\mathbf{e}_i), i \in J \setminus U \\ & \text{or } (j, \mathbf{v}) = (0, \mathbf{0}), i \in U, \\ 0, & \text{otherwise,} \end{cases} \quad (2.7)$$

where $\mathbf{0}$ is a null vector. Note that $\mathbf{X}^{(0U)}$ represents the increment when the queues with indices in U are empty. Then, $\{\mathbf{Z}_\ell\}$ can be obtained as

$$\mathbf{Z}_{\ell+1} = \mathbf{Z}_\ell + \sum_{U \in \mathcal{K}} \left(\mathbf{X}_\ell^{(+U)} 1(\mathbf{Z}_\ell \in \mathcal{S}_{+U}) + \mathbf{X}_\ell^{(0U)} 1(\mathbf{Z}_\ell \in \mathcal{S}_{0U}) \right), \quad (2.8)$$

where $\mathbf{X}_\ell^{(+U)}$ and $\mathbf{X}_\ell^{(0U)}$ are independent copies of $\mathbf{X}^{(+U)}$ and $\mathbf{X}^{(0U)}$, respectively, and $1(\cdot)$ is the indicator function. This $\{\mathbf{Z}_\ell\}$ is referred to as a reflecting random walk for the JSQ.

By the stability condition (2.1), $\{\mathbf{Z}_\ell\}$ has a stationary distribution. Denote a random vector subject to this distribution by $\mathbf{Z} \equiv (M, \mathbf{Y})$. Then, (2.8) yields

$$\mathbf{Z} \stackrel{d}{=} \mathbf{Z} + \sum_{U \in \mathcal{K}} \left(\mathbf{X}^{(+U)} 1(\mathbf{Z} \in \mathcal{S}_{+U}) + \mathbf{X}^{(0U)} 1(\mathbf{Z} \in \mathcal{S}_{0U}) \right), \quad (2.9)$$

where $\stackrel{d}{=}$ stands for equality in distribution, and $\mathbf{X}^{(+U)}$ and $\mathbf{X}^{(0U)}$ are assumed to be independent of \mathbf{Z} . The stationary equation (2.9) plays a key roll in our arguments.

Let

$$\mathcal{H} = \{\mathbf{h} = (h_1, h_2, \dots, h_k) \in \mathbb{Z}_+^k; \exists i \in J, h_i = 0\}.$$

Obviously, \mathcal{H} is a state space for \mathbf{Y} and $\mathcal{S} = \mathbb{Z}_+ \times \mathcal{H}$. We are ready to present our main result of this paper which will be proved in Section 3 using results in Section 4 and Appendices.

Theorem 2.1 For the M/M -JSQ with k parallel queues satisfying the stability condition (2.1), we have for each $\mathbf{h} \in \mathcal{H}$,

$$\lim_{n \rightarrow \infty} \rho^{-kn} \mathbb{P}(M = n, \mathbf{Y} = \mathbf{h}) = c_{\mathbf{h}}, \quad (2.10)$$

where $c_{\mathbf{h}}$ is a positive constant.

For two parallel queues with homogeneous servers, this theorem was firstly obtained by Kingman [3] by using analytic functions. It is also known for two parallel queues with heterogeneous servers (e.g, see [15]). Similar results were obtained for two parallel queues under more general setting (e.g., see [2, 6, 8, 13, 15] and references in those papers). Many of them use the QBD processes and their limiting behaviors (e.g., see [10]).

Corollary 2.1 Under the same assumptions of Theorem 2.1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(M = n) = \log \rho^k. \quad (2.11)$$

This corollary is not immediate from Theorem 2.1 because (2.10) only implies that the limit infimum of $\frac{1}{n} \log \mathbb{P}(M = n)$ is lower bounded by $\log \rho^k$. We will prove it in Appendix E.

3 Proof of Theorem 2.1

For the proof of Theorem 2.1, we use two formulations, the QBD process and the reflecting random walk for the JSQ. We first discuss basic results on the exact asymptotic for the QBD process (see Proposition 3.1). We next consider the convergence domain for the moment generating function of stationary distribution. For this, we prepare some lemmas. The last lemma among them has a key roll in our arguments, which will be proved in Section 4. Finally, we prove Theorem 2.1 in Section 3.3.

3.1 QBD process and sufficient conditions for geometric tail decay

We first present the tail asymptotic result for the QBD process known in the literature [6, 10]. For this, we use some matrices. For $i = 0, \pm 1$ and $\ell = 0, 1, 2, \dots$, define infinite dimensional matrices B_0, A_i as

$$\begin{aligned} [B_0]_{\mathbf{h}, \mathbf{h}'} &= \mathbb{P}(\mathbf{Z}_{\ell+1} = (0, \mathbf{h}') | \mathbf{Z}_\ell = (0, \mathbf{h})), & \mathbf{h}, \mathbf{h}' \in \mathcal{H}, \\ [A_i]_{\mathbf{h}, \mathbf{h}'} &= \mathbb{P}(\mathbf{Z}_{\ell+1} = (n+i, \mathbf{h}') | \mathbf{Z}_\ell = (n, \mathbf{h})), & \mathbf{h}, \mathbf{h}' \in \mathcal{H}, n \geq 1. \end{aligned}$$

Then, the QBD process $\{\mathbf{Z}_\ell\}$ has the following transition probability matrix.

$$P = \begin{pmatrix} B_0 & A_{+1} & & & \\ A_{-1} & A_0 & A_{+1} & & \\ & A_{-1} & A_0 & A_{+1} & \\ & & \ddots & \ddots & \ddots \end{pmatrix}.$$

We assume the stability condition (2.1), and therefore the stationary distribution exists. We denote it by row vector $\boldsymbol{\pi}$.

For the QBD process, it is important to distinguish level from background state. For this, we partition the stationary vector $\boldsymbol{\pi}$ as $(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \dots)$ according to level. That is,

$$[\boldsymbol{\pi}_n]_{\mathbf{h}} = \mathbb{P}(M = n, \mathbf{Y} = \mathbf{h}),$$

for $n \in \mathbb{Z}_+$ and $\mathbf{h} \in \mathcal{H}$, where $[\boldsymbol{\pi}_n]_{\mathbf{h}}$ denotes the \mathbf{h} -th entry of $\boldsymbol{\pi}_n$. As is well known (e.g., [5, 11]), the stationary distribution is known to have the following matrix geometric form:

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 R^n, \quad n \geq 1, \tag{3.1}$$

where R is the minimal nonnegative solution of the following equation:

$$R = A_{+1} + RA_0 + R^2 A_{-1}. \tag{3.2}$$

When the size of R is finite, we can see that the tail decay rate of (3.1) is obtained as the maximal eigenvalue of R . Otherwise, this is not always true. Thus, we need certain extra conditions here. Such conditions were firstly obtained in [15], and refined and generalized in [10]. The following result for the geometric tail decay are the specialization of the results in [10] to the QBD process (see Theorem 2.1 of [6]).

Proposition 3.1 Assume that $A \equiv A_{-1} + A_0 + A_{+1}$ is irreducible and aperiodic, and that the Markov additive process generated by $\{A_\ell; \ell = 0, \pm 1\}$ is 1-arithmetic. If there exist $\alpha > 1$ and positive vectors \mathbf{x} and \mathbf{y} such that

$$\mathbf{x}A_*(\alpha) = \mathbf{x}, \quad A_*(\alpha)\mathbf{y} = \mathbf{y}, \quad (3.3)$$

$$\mathbf{x}\mathbf{y} < \infty, \quad (3.4)$$

where $A_*(z) = z^{-1}A_{-1} + A_0 + zA_{+1}$ for $z \neq 0$, then R has left and right eigenvectors \mathbf{x} and $\mathbf{r} \equiv (I - A_0 - RA_{-1} - \alpha^{-1}A_{-1})\mathbf{y}$, respectively, with eigenvalue α^{-1} . Furthermore, if

$$\pi_0\mathbf{y} < \infty, \quad (3.5)$$

then we have the following geometric tail asymptotics for the stationary distribution:

$$\lim_{n \rightarrow \infty} \alpha^n \pi_n = \frac{\pi_0 \mathbf{r}}{\mathbf{x} \mathbf{r}} \mathbf{x}. \quad (3.6)$$

Remark 3.1 Both irreducibility and aperiodicity of A are easy to verify for our queueing model. Furthermore, the 1-arithmetic property is directly verified, that is, for each $\mathbf{h} \in \mathcal{H}$, the greatest common divisor of

$$\{\ell_1 + \ell_2 + \cdots + \ell_i; [A_{\ell_1}]_{\mathbf{h}, \mathbf{h}_1} [A_{\ell_2}]_{\mathbf{h}_1, \mathbf{h}_2} \times \cdots \times [A_{\ell_i}]_{\mathbf{h}_{i-1}, \mathbf{h}}, \\ \text{where } i \geq 1, \ell_m = 0, \pm 1, m = 1, 2, \dots, i, \mathbf{h}_n \in \mathcal{H}, n = 1, 2, \dots, i-1\}$$

is shown to be one.

It is not very hard to find positive vectors \mathbf{x} and \mathbf{y} satisfying conditions (3.3) and (3.4). We first derive the α and the right invariant vector \mathbf{y} in (3.3)

Lemma 3.1 Let $y_{\mathbf{h}} = \rho^{-\mathbf{h}1}$ for $\mathbf{h} \in \mathcal{H}$. Then, $\alpha = \rho^{-k}$ and $\mathbf{y} = (y_{\mathbf{h}}; \mathbf{h} \in \mathcal{H})$ satisfy $A_*(\alpha)\mathbf{y} = \mathbf{y}$.

This lemma is proved in Appendix A. We next consider left invariant vector \mathbf{x} and (3.4). Let $\Delta_{\mathbf{y}}$ be the diagonal matrix whose \mathbf{h} -th diagonal element is $y_{\mathbf{h}}$ for $\mathbf{h} \in \mathcal{H}$ and the other entires are 0. Since \mathbf{y} is the right invariant vector of $A_*(\rho^{-k})$, $\Delta_{\mathbf{y}}^{-1}A_*(\rho^{-k})\Delta_{\mathbf{y}}$ is a stochastic matrix. We have the following lemma, which is proved in Appendix B.

Lemma 3.2 The stochastic matrix $\Delta_{\mathbf{y}}^{-1}A_*(\rho^{-k})\Delta_{\mathbf{y}}$ is positive recurrent.

Using these lemmas, we will verify the conditions (3.3) and (3.4) in Section 3.3.

We finally consider the condition (3.5). However, we need much effort to check condition (3.5) since it includes the unknown vector π_0 . For this, we will consider a convergence domain of the moment generating function for the JSQ in Section 3.2. In the next subsection, we only present results on the domain as lemmas, and prove them in Appendices C and D and in Section 4.

3.2 Stationary inequality for moment generating functions

For $\boldsymbol{\theta} \equiv (\theta_0, \theta_1, \theta_2, \dots, \theta_k) \in \mathbb{R}^{k+1}$, where \mathbb{R} is the set of all real numbers, let $\varphi(\boldsymbol{\theta})$ denote the moment generating function of the random vector $\mathbf{Z} = (M, \mathbf{Y})$ in (2.9), that is,

$$\varphi(\boldsymbol{\theta}) = \mathbb{E}(e^{\boldsymbol{\theta}\mathbf{Z}}), \quad (3.7)$$

where \mathbf{ab} denotes the inner product of vectors \mathbf{a} and \mathbf{b} . We are interested in the convergence domain of φ which is denoted by \mathcal{D} , i.e.,

$$\mathcal{D} = \{\boldsymbol{\theta} \in \mathbb{R}^{k+1}; \varphi(\boldsymbol{\theta}) < \infty\}. \quad (3.8)$$

In what follows, we will study the domain \mathcal{D} by using the stationary equation for the moment generating function. To this end, we introduce some notations. For $U \in \mathcal{K}$, let γ_{+U} and γ_{0U} be the moment generating functions of the random vectors $\mathbf{X}^{(+U)}$ and $\mathbf{X}^{(0U)}$, respectively, that is,

$$\gamma_{+U}(\boldsymbol{\theta}) = \mathbb{E}(e^{\boldsymbol{\theta}\mathbf{X}^{(+U)}}), \quad \gamma_{0U}(\boldsymbol{\theta}) = \mathbb{E}(e^{\boldsymbol{\theta}\mathbf{X}^{(0U)}}), \quad \boldsymbol{\theta} \in \mathbb{R}^{k+1}.$$

From equations (2.4)–(2.7), we have

$$\gamma_{+U}(\boldsymbol{\theta}) = \begin{cases} \lambda e^{\theta_0 - \sum_{j \in J \setminus U} \theta_j} + \sum_{j \in J \setminus U} \mu_j e^{-\theta_j} + \mu_i e^{-\theta_0 + \sum_{j \in J \setminus U} \theta_j} & \text{for } |U| = 1, i \in U, \\ \sum_{i \in U} \frac{1}{|U|} \lambda e^{\theta_i} + \sum_{i \in J \setminus U} \mu_i e^{-\theta_i} + \sum_{i \in U} \mu_i e^{-\theta_0 + \sum_{j \in J \setminus \{i\}} \theta_j} & \text{for } |U| \geq 2, \end{cases} \quad (3.9)$$

$$\gamma_{0U}(\boldsymbol{\theta}) = \begin{cases} \lambda e^{\theta_0 - \sum_{j \in J \setminus U} \theta_j} + \sum_{j \in J \setminus U} \mu_j e^{-\theta_j} + \mu_i & \text{for } |U| = 1, i \in U, \\ \sum_{i \in U} \frac{1}{|U|} \lambda e^{\theta_i} + \sum_{i \in J \setminus U} \mu_i e^{-\theta_i} + \sum_{i \in U} \mu_i & \text{for } |U| \geq 2. \end{cases} \quad (3.10)$$

We further define two moment generating functions φ_{+U} and φ_{0U} as follows:

$$\varphi_{+U}(\boldsymbol{\theta}) = \mathbb{E}(e^{\boldsymbol{\theta}\mathbf{Z}} 1(\mathbf{Z} \in \mathcal{S}_{+U})), \quad (3.11)$$

$$\varphi_{0U}(\boldsymbol{\theta}) = \mathbb{E}(e^{\boldsymbol{\theta}\mathbf{Z}} 1(\mathbf{Z} \in \mathcal{S}_{0U})), \quad (3.12)$$

for $\boldsymbol{\theta} \in \mathbb{R}^{k+1}$.

Remark 3.2 For each $U \in \mathcal{K}$ and $\boldsymbol{\theta} \in \mathbb{R}^{k+1}$, $\varphi_{+U}(\boldsymbol{\theta})$ does not depend on the parameter θ_i for $i \in U$ since the expectation in (3.11) is taken over the event $\{\mathbf{Z} \in \mathcal{S}_{+U}\}$, i.e., $Y_i = 0$ for all $i \in U$. Similarly, $\varphi_{0U}(\boldsymbol{\theta})$ does not depend on the parameters θ_0 and θ_i for $i \in U$.

From (3.9) and (3.10), for all $U \in \mathcal{K}$, it is easy to see that $\gamma_{+U}(\boldsymbol{\theta})$ and $\gamma_{0U}(\boldsymbol{\theta})$ are finite for all $\boldsymbol{\theta} \in \mathbb{R}^{k+1}$. Thus, from (2.9), we have a stationary equation of moment generating function as long as $\varphi(\boldsymbol{\theta})$ is finite.

$$\varphi(\boldsymbol{\theta}) = \sum_{U \in \mathcal{K}} (\gamma_{+U}(\boldsymbol{\theta}) \varphi_{+U}(\boldsymbol{\theta}) + \gamma_{0U}(\boldsymbol{\theta}) \varphi_{0U}(\boldsymbol{\theta})). \quad (3.13)$$

Furthermore, partitioning $\varphi(\boldsymbol{\theta})$ concerning each boundary, we have the following decomposition:

$$\varphi(\boldsymbol{\theta}) = \sum_{U \in \mathcal{K}} (\varphi_{+U}(\boldsymbol{\theta}) + \varphi_{0U}(\boldsymbol{\theta})). \quad (3.14)$$

From (3.13) and (3.14), we get

$$\sum_{U \in \mathcal{K}} ((1 - \gamma_{+U}(\boldsymbol{\theta}))\varphi_{+U}(\boldsymbol{\theta}) + (1 - \gamma_{0U}(\boldsymbol{\theta}))\varphi_{0U}(\boldsymbol{\theta})) = 0. \quad (3.15)$$

The stationary equation (3.15) holds at least $\boldsymbol{\theta} \leq \mathbf{0}$.

We consider the convergence domain of the moment generating function $\varphi(\boldsymbol{\theta})$ for $\boldsymbol{\theta} > \mathbf{0}$. We first prove that the distribution of \mathbf{Z} has a light tail. For this, we will use the idea of Foley and McDonald [2], in which they obtain a similar result for $k = 2$.

Lemma 3.3 Under the stability condition (2.1), there exists an $\epsilon > 0$ such that $\mathbb{E}(e^{\epsilon \sum_{i=1}^k L_i}) < \infty$, where $L_i = M + Y_i$ for $i \in J$, and therefore $\varphi(0, \epsilon \mathbf{1}) < \infty$.

The proof of this lemma is deferred to Appendix C. From this lemma, we find a confirmed region for the convergence domain of the moment generating function φ in the positive orthant. Starting with this region, we will expand the confirmed region of the convergence domain \mathcal{D} . To this end, we consider the stationary equation (3.15). However, we only know that the stationary equation (3.15) holds with $\boldsymbol{\theta} \leq \mathbf{0}$. Thus, we can not use the stationary equation (3.15) directly. Instead of doing so, we derive inequalities on the stationary distribution. For this, we recall that \mathcal{K} is the set of all subsets of J except for empty set.

Lemma 3.4 For fix $\boldsymbol{\theta} \in \mathbb{R}^{k+1}$, we have the following results.

(a) If $\varphi_{0U}(\boldsymbol{\theta}) < \infty$ and $1 - \gamma_{+U}(\boldsymbol{\theta}) > 0$ for all $U \in \mathcal{K}$, then we have

$$\sum_{U \in \mathcal{K}} (1 - \gamma_{+U}(\boldsymbol{\theta}))\varphi_{+U}(\boldsymbol{\theta}) \leq \sum_{U \in \mathcal{K}} (\gamma_{0U}(\boldsymbol{\theta}) - 1)\varphi_{0U}(\boldsymbol{\theta}) < \infty. \quad (3.16)$$

(b) Let \mathcal{A} be a subset \mathcal{K} . If

$$\gamma_{+U}(\boldsymbol{\theta}) < 1, \gamma_{0U}(\boldsymbol{\theta}) < 1, \quad U \in \mathcal{K} \setminus \mathcal{A}, \quad (3.17)$$

$$\varphi_{+U'}(\boldsymbol{\theta}) < \infty, \varphi_{0U'}(\boldsymbol{\theta}) < \infty, \quad U' \in \mathcal{A}, \quad (3.18)$$

then we have

$$\begin{aligned} & \sum_{U \in \mathcal{K} \setminus \mathcal{A}} ((1 - \gamma_{+U}(\boldsymbol{\theta}))\varphi_{+U}(\boldsymbol{\theta}) + (1 - \gamma_{0U}(\boldsymbol{\theta}))\varphi_{0U}(\boldsymbol{\theta})) \\ & \leq \sum_{U' \in \mathcal{A}} ((\gamma_{+U'}(\boldsymbol{\theta}) - 1)\varphi_{+U'}(\boldsymbol{\theta}) + (\gamma_{0U'}(\boldsymbol{\theta}) - 1)\varphi_{0U'}(\boldsymbol{\theta})) < \infty. \end{aligned} \quad (3.19)$$

(c) If all the conditions in either (a) or (b) hold, then $\varphi(\boldsymbol{\theta}) < \infty$.

Remark 3.3 If $\varphi_{+U}(\boldsymbol{\theta})$ and $\varphi_{0U}(\boldsymbol{\theta})$ are finite for all $U \in \mathcal{K}$, then the stationary equation (3.13) is satisfied, and we have (3.16) and (3.19) with equalities. The important claim of this lemma is that (3.16) and (3.19) hold even if the finiteness of the left hand sides are unknown.

This lemma is an adaptation of Lemma 6.4 in [9], but we prove it in Appendix D for this paper to be selfcontained. Using these inequalities, we obtain a part of domain \mathcal{D} to be sufficient for our purpose. We prove the following lemma in Section 4.

Lemma 3.5 For $(\eta_0, \eta_1) \in \mathbb{R}^2$, if

$$\eta_0 < \log \rho^{-k}, \quad \eta_1 < \frac{1}{k-1} \log \rho^{-k} = \log \rho^{-1} + \frac{1}{k-1} \log \rho^{-1}, \quad (3.20)$$

then we have $\varphi(\eta_0, \eta_1 \mathbf{1}) < \infty$.

This fact will be used to verify the condition (3.5). We are now ready to prove Theorem 2.1 using Proposition 3.1.

3.3 Verifying the sufficient conditions in Proposition 3.1

From Lemma 3.1, we already obtain $\alpha = \rho^{-k}$ and $\mathbf{y} = (y_{\mathbf{h}})$, where $y_{\mathbf{h}} = \rho^{-\mathbf{h}\mathbf{1}}$ for $\mathbf{h} \in \mathcal{H}$. So, we need to check the conditions (3.3), (3.4) and (3.5). From Lemma 3.2, the stationary distribution of $\Delta_{\mathbf{y}}^{-1} A_*(\rho^{-k}) \Delta_{\mathbf{y}}$ exists, and denote it by $\boldsymbol{\nu}$, that is,

$$\boldsymbol{\nu} \Delta_{\mathbf{y}}^{-1} A_*(\rho^{-k}) \Delta_{\mathbf{y}} = \boldsymbol{\nu}.$$

Thus, the condition (3.3) holds with $\mathbf{x} = \boldsymbol{\nu} \Delta_{\mathbf{y}}^{-1}$. In addition, we have

$$\mathbf{x} \mathbf{y} = \boldsymbol{\nu} \Delta_{\mathbf{y}}^{-1} \mathbf{y} = 1.$$

Hence, the condition (3.4) is satisfied.

We finally verify the condition (3.5). We have the following equation.

$$\boldsymbol{\pi}_0 \mathbf{y} = \sum_{\mathbf{h} \in \mathcal{H}} [\boldsymbol{\pi}_0]_{\mathbf{h}} \rho^{-\mathbf{h}\mathbf{1}} = \sum_{\mathbf{h} \in \mathcal{H}} [\boldsymbol{\pi}_0]_{\mathbf{h}} e^{\mathbf{h}\mathbf{1} \log \rho^{-1}}. \quad (3.21)$$

Since φ is the moment generating function of stationary distribution, for $(\eta_0, \eta_1) \in \mathbb{R}^2$,

$$\varphi(\eta_0, \eta_1 \mathbf{1}) = \sum_{n=0}^{\infty} \sum_{\mathbf{h} \in \mathcal{H}} [\boldsymbol{\pi}_n]_{\mathbf{h}} e^{\eta_0 n + \eta_1 \mathbf{h}\mathbf{1}}. \quad (3.22)$$

From (3.21) and (3.22), we have

$$\boldsymbol{\pi}_0 \mathbf{y} \leq \varphi(0, \log \rho^{-1} \mathbf{1}). \quad (3.23)$$

By Lemma 3.5, we have $\varphi(0, \log \rho^{-1} \mathbf{1}) < \infty$, and the left hand side of (3.23) is finite. Thus, the conditions in Proposition 3.1 are satisfied. This completes the proof of Theorem 2.1.

4 The proof of Lemma 3.5

The main object of this section is to prove Lemma 3.5. Let

$$\mathcal{D}^{(2)} = \{(\eta_0, \eta_1) \in \mathbb{R}^2; (\eta_0, \eta_1 \mathbf{1}) \in \mathcal{D}\}.$$

That is, $\mathcal{D}^{(2)}$ is the projection \mathcal{D} to the two dimensional hyper plane. Then, for $(\eta_0, \eta_1) \in \mathbb{R}^2$, $\varphi(\eta_0, \eta_1 \mathbf{1}) < \infty$ is equivalent to $(\eta_0, \eta_1) \in \mathcal{D}^{(2)}$. Thus, we will show that $(\eta_0, \eta_1) \in \mathcal{D}^{(2)}$ under the condition (3.20). We iteratively find a sequence of vectors $\zeta_1, \zeta_2, \dots, \zeta_m$ for some $m \geq 1$ such that $\zeta_\ell \in \mathcal{D}^{(2)}$ for $1 \leq \ell \leq m$ and $(\eta_0, \eta_1) \leq \zeta_\ell$. Our approach is similar to [4, 9]. One may think that we can still use the two dimensional iteration. However, things are not so simple because of the $k+1$ -dimensional nature of the stationary distribution. In particular, we must consider the $k+1$ dimensional moment generating function when we expand the confirmed region in the direction of η_1 -axis. To overcome this issue, we prepare some technical lemmas.

4.1 The first step for expanding the confirmed region

We consider to iteratively expand the confirmed region for $\varphi(\boldsymbol{\theta}) < \infty$ by using Lemma 3.4. For this, it is important to suitably choose \mathcal{A} for Lemma 3.4. For each $V \in \mathcal{K}$, let

$$\mathcal{G}(V) = \{U \in \mathcal{K}; V \cap U \neq \emptyset\},$$

and denote $\mathcal{K} \setminus \mathcal{G}(V)$ by $\mathcal{G}^c(V)$. It follows from Lemma 3.4 with $\mathcal{A} = \mathcal{G}(V)$ that

$$\gamma_{+U}(\boldsymbol{\theta}) < 1, \gamma_{0U}(\boldsymbol{\theta}) < 1, \quad U \in \mathcal{G}^c(V), \quad (4.1)$$

$$\varphi_{+U}(\boldsymbol{\theta}) < \infty, \varphi_{0U}(\boldsymbol{\theta}) < \infty, \quad U \in \mathcal{G}(V), \quad (4.2)$$

imply that $\varphi(\boldsymbol{\theta}) < \infty$.

We next consider to minimize the number of the conditions in (4.2) to be verified by restricting to the case that $|V| = 1$. To this end, we introduce some notations. For each $i = 0, 1$, let $\{\zeta_{i,j}; i = 0, 1, j \in J\}$ be a $2 \times k$ matrix. For $V \in \mathcal{K}$, let $\zeta(V)$ be the $k+1$ -dimensional vector whose components are give by

$$[\zeta(V)]_i = \begin{cases} \zeta_{0,|V|} & \text{for } i = 0, \\ \zeta_{1,|V|} & \text{for } i \in V, \\ 0 & \text{for } i \in J \setminus V. \end{cases} \quad (4.3)$$

Then, we have the following lemma.

Lemma 4.1 For each $i = 0, 1$, assume that $\zeta_{i,j}$ is nonincreasing in $j \in J$ and

(C1) $\gamma_{+U}(\zeta(V)) < 1$ and $\gamma_{0U}(\zeta(V)) < 1$ for each $V \in \mathcal{K} \setminus \{J\}$, that is, $V \in \mathcal{K}$ satisfying $|V| \leq k-1$, and $U \in \mathcal{G}^c(V)$.

Then, the condition:

(C2) $\varphi_{+U}(\zeta(V)) < \infty$ for each $V \in \mathcal{K} \setminus \{J\}$ satisfying $|V| = 1$ and each $U \in \mathcal{G}(V)$

implies that

$$(\zeta_{0,1}, \zeta_{1,k-1}) \in \mathcal{D}^{(2)}. \quad (4.4)$$

Remark 4.1 For $V \in \mathcal{K} \setminus \{J\}$ such that $|V| = 1$, $U \in \mathcal{G}(V)$ if and only if $U = V$, and therefore, for this U , $\varphi_{+U}(\zeta(V))$ does not depend on the elements of $\zeta(V)$ except for first entry, i.e., $\zeta_{0,1}$.

PROOF. For $(\zeta_{0,k-1}, \zeta_{1,k-1}) \in \mathbb{R}^2$, from (3.14), we have

$$\varphi(\zeta_{0,k-1}, \zeta_{1,k-1} \mathbf{1}) = \sum_{U \in \mathcal{K}} (\varphi_{0U}(\zeta_{0,k-1}, \zeta_{1,k-1} \mathbf{1}) + \varphi_{+U}(\zeta_{0,k-1}, \zeta_{1,k-1} \mathbf{1})). \quad (4.5)$$

Since (C1) holds for $V \in \mathcal{K} \setminus \{J\}$, (4.4) is obtained if we verify

$$\varphi_{+U}(\zeta(V)) < \infty, \quad V \in \mathcal{K} \setminus \{J\}, U \in \mathcal{G}(V), \quad (4.6)$$

by Lemma 3.4.

Assume the condition (C2). We inductively verify (4.6) on the value of $|V|$, where $1 \leq |V| \leq k-1$. (4.6) holds for $|V| = 1$ by (C2). For a fixed ℓ , where $1 < \ell < k-1$, we assume that, for $V \in \mathcal{K}$ satisfying $1 \leq |V| \leq \ell$,

$$\varphi_{+U}(\zeta(V)) < \infty, \quad \varphi_{0U}(\zeta(V)) < \infty, \quad U \in \mathcal{G}(V). \quad (4.7)$$

If we can show that (4.7) for $V \in \mathcal{K}$ such that $|V| = \ell+1$, then the induction is completed, and therefore we have (4.6).

Arbitrarily choose $U' \in \mathcal{G}(V')$ for $V' \in \mathcal{K}$ satisfying $|V'| = \ell+1$. We recall that the expectations in $\varphi_{+U'}$ and $\varphi_{0U'}$ are taken over the events $\{Y_i = 0; i \in U'\}$ and $\{M = 0\} \cap \{Y_i = 0; i \in U'\}$, respectively. Hence, we have

$$\varphi_{+U'}(\zeta(V')) \leq \varphi_{+U'}(\zeta(V' \setminus U')), \quad \varphi_{0U'}(\zeta(V')) \leq \varphi_{0U'}(\zeta(V' \setminus U')), \quad (4.8)$$

since $\zeta_{0,|V'|} \leq \zeta_{0,|V' \setminus U'|}$ and $\zeta_{1,|V'|} \leq \zeta_{1,|V' \setminus U'|}$ by the nonincreasing assumption (see also (4.3)). We note that $|V' \setminus U'| \leq \ell$ for $U' \in \mathcal{G}(V')$ since $U' \cap V' \neq \emptyset$. Thus, from the induction assumption (4.7) for $|V| \leq \ell$, (4.7) is satisfied for $|V| = \ell+1$. This completes the proof of the lemma. \square

From this lemma, we have the following fact, which will be used to expand the confirmed region.

Lemma 4.2 If $(\eta_0, 0) \in \mathcal{D}^{(2)}$ for $\eta_0 > 0$, then $(\eta_0, \frac{1}{k-1}\eta_0) \in \mathcal{D}^{(2)}$.

PROOF. We will use Lemma 4.1. For this, let $\zeta_{0,j} = \eta_0$ and $\zeta_{1,j} = \frac{1}{j}\eta_0$. Then, for $i = 0, 1$, $\zeta_{i,j}$ is nonincreasing in j . Moreover, from Remark 4.1, we have (C2) from our assumption. If we can verify (C1), then for $\zeta_{0,j} = \eta_0$ and $\zeta_{1,j} = \frac{1}{j}\eta_0$, all conditions of Lemma 4.1 are satisfied, and therefore, $(\eta_0, \frac{1}{k-1}\eta_0) \in \mathcal{D}^{(2)}$.

In what follows, we check the condition (C1). For this, we consider the case $U \in \mathcal{G}^c(V)$. We recall that $V \cap U = \emptyset$. Thus, for $j \in J$, $\zeta_{0,j} = \eta_0$ and $\zeta_{1,j} = \frac{1}{j}\eta_0$,

$$\begin{aligned}
[\zeta(V)]_i &= ([\zeta(V)]_i 1(i \in V) + [\zeta(V)]_i 1(i \in J \setminus V)) = 0, & \forall i \in U, \\
\sum_{i \in J \setminus U} [\zeta(V)]_i &= \sum_{i \in J \setminus U} ([\zeta(V)]_i 1(i \in V) + [\zeta(V)]_i 1(i \in J \setminus V)) = \eta_0, & |U| = 1, \\
\sum_{j \in J \setminus \{i\}} [\zeta(V)]_j &= \sum_{j \in J \setminus \{i\}} ([\zeta(V)]_j 1(j \in V) + [\zeta(V)]_j 1(j \in J \setminus V)) = \eta_0, & \forall i \in U, \\
\sum_{i \in J \setminus U} \mu_i e^{-[\zeta(V)]_i} &= \sum_{i \in J \setminus U} \mu_i (e^{-[\zeta(V)]_i} 1(i \in V) + e^{-[\zeta(V)]_i} 1(i \notin V)) \\
&= \sum_{i \in V} \mu_i e^{-\frac{1}{|V|}\eta_0} + \sum_{i \in J \setminus (U \cup V)} \mu_i.
\end{aligned}$$

For any i satisfying $U = \{i\}$, that is $|U| = 1$, substituting these equations into (3.9) and (3.10), we have

$$\begin{aligned}
\gamma_{+U}(\zeta(V)) &= \lambda e^{\zeta_{0,|V|} - \sum_{j \in J \setminus U} [\zeta(V)]_j} + \sum_{j \in J \setminus U} \mu_j e^{-[\zeta(V)]_j} + \mu_i e^{-\zeta_{0,|V|} + \sum_{j \in J \setminus \{i\}} [\zeta(V)]_j} \\
&= \lambda + \sum_{j \in V} \mu_j e^{-\frac{1}{|V|}\eta_0} + \sum_{j \in J \setminus (U \cup V)} \mu_j + \mu_i \\
&< \lambda + \sum_{j \in V} \mu_j + \sum_{j \in J \setminus (U \cup V)} \mu_j + \mu_i \\
&= \lambda + \sum_{j=1}^k \mu_j \\
&= 1, \\
\gamma_{0U}(\zeta(V)) &= \lambda e^{\zeta_{0,|V|} - \sum_{j \in J \setminus U} [\zeta(V)]_j} + \sum_{j \in J \setminus U} \mu_j e^{-[\zeta(V)]_j} + \mu_i \\
&= \lambda + \sum_{j \in V} \mu_j e^{-\frac{1}{|V|}\eta_0} + \sum_{j \in J \setminus (U \cup V)} \mu_j + \mu_i \\
&< 1,
\end{aligned}$$

where we used the assumption (2.2). On the other hand, for $|U| \geq 2$, we have

$$\begin{aligned}
\gamma_{+U}(\zeta(V)) &= \sum_{i \in U} \frac{1}{|U|} \lambda e^{[\zeta(V)]_i} + \sum_{i \in J \setminus U} \mu_i e^{-[\zeta(V)]_i} + \sum_{i \in U} \mu_i e^{-\zeta_{0,|V|} + \sum_{j \in J \setminus \{i\}} [\zeta(V)]_j} \\
&= \lambda + \sum_{i \in V} \mu_i e^{-\frac{1}{|V|}\eta_0} + \sum_{i \in J \setminus (U \cup V)} \mu_i + \sum_{i \in U} \mu_i \\
&< 1. \\
\gamma_{0U}(\zeta(V)) &= \sum_{i \in U} \frac{1}{|U|} \lambda e^{[\zeta(V)]_i} + \sum_{i \in J \setminus U} \mu_i e^{-[\zeta(V)]_i} + \sum_{i \in U} \mu_i \\
&= \lambda + \sum_{i \in V} \mu_i e^{-\frac{1}{|V|}\eta_0} + \sum_{i \in J \setminus (U \cup V)} \mu_i + \sum_{i \in U} \mu_i \\
&< 1.
\end{aligned}$$

Hence, for each fixed V , we have (C1). This completes the proof. \square

4.2 Iterations for expansion

We next consider two dimensional marginals of the moment generating functions (3.9) and (3.10). For each $U \in \mathcal{K}$ and $(\eta_1, \eta_2) \in \mathbb{R}^2$, we define the moment generating functions $\gamma_{+U}^{(2)}$ and $\gamma_{0U}^{(2)}$ as

$$\gamma_{+U}^{(2)}(\eta_0, \eta_1) = \gamma_{+U}(\eta_0, \eta_1 \mathbf{1}), \quad \gamma_{0U}^{(2)}(\eta_0, \eta_1) = \gamma_{0U}(\eta_0, \eta_1 \mathbf{1}).$$

For $U \in \mathcal{K}$. let

$$\begin{aligned} \Gamma_{+U}^{(2)} &= \{\boldsymbol{\eta} = (\eta_0, \eta_1) \in \mathbb{R}^2; \gamma_{+U}^{(2)}(\boldsymbol{\eta}) \leq 1\}, \\ \partial\Gamma_{+U}^{(2)} &= \{\boldsymbol{\eta} = (\eta_0, \eta_1) \in \mathbb{R}^2; \gamma_{+U}^{(2)}(\boldsymbol{\eta}) = 1\}. \end{aligned}$$

Then, we have the following facts.

Proposition 4.1 Under the assumption of Theorem 2.1, we have the following properties.

- (i) $\Gamma_{+U}^{(2)}$ is a convex set in \mathbb{R}^2 for $U \in \mathcal{K}$.
- (ii) $\{\eta_i \in \mathbb{R}; (\eta_0, \eta_1) \in \partial\Gamma_{+U}^{(2)}\}$ is bounded for each fixed η_{1-i} and $i = 0, 1$ and $U \in \mathcal{K}$ such that $|U| = 1$.
- (iii) $(0, \eta_1) \in \partial\Gamma_{+U}^{(2)}$ for some $\eta_1 > 0$. and $U \in \mathcal{K}$ satisfying $|U| = 1$.
- (iv) For $U \in \mathcal{K}$ such that $U \geq 2$, $\{\eta_1 \in \mathbb{R}; (\eta_0, \eta_1) \in \partial\Gamma_{+U}^{(2)}, \exists \eta_0 \in \mathbb{R}\}$ is bounded from above.
- (v) All the curves $\partial\Gamma_{+U}^{(2)}$ for $U \in \mathcal{K}$ intersect at $(0, 0)$ and $(\log \rho^{-k}, \log \rho^{-1})$.

We obviously see (i)–(iv). For example, (i) is obtained because $\gamma_{+U}^{(2)}$ is a convex function. Furthermore, (v) is obtained by letting $\theta_0 = \log \rho^{-k}$ and $\theta_1 = \theta_2 = \dots = \theta_k = \log \rho^{-1}$ in (3.9). So far, we omit a detailed proof of this proposition. In Figure 2, for $k = 2$, we depict the convex curve $\partial\Gamma_{+U}^{(2)}$ and the region where the condition (3.20) holds.

Using Lemma 4.2, we iteratively find nondecreasing point such that the moment generating function $\varphi(\eta_0, \eta_1 \mathbf{1})$ is finite. We illustrate our iteration in Figure 2. For this, we recall that there exists $\bar{\eta}_1^{(0)} > 0$ such that $\varphi(0, \bar{\eta}_1^{(0)} \mathbf{1})$ is finite by Lemma 3.3. Let $\bar{\boldsymbol{\eta}}^{(0)} = (0, \bar{\eta}_1^{(0)})$ and for $\ell = 1, 2, \dots$,

$$\boldsymbol{\eta}^{(\ell)} = (\eta_0^{(\ell)}, \eta_1^{(\ell)}) \equiv \arg \sup_{(\eta_0, \eta_1)} \{\eta_0; \gamma_{+U}^{(2)}(\boldsymbol{\eta}) < 1, \forall U \in \mathcal{K}, \eta_1 \leq \bar{\eta}_1^{(\ell-1)}\}, \quad (4.9)$$

$$\bar{\boldsymbol{\eta}}^{(\ell)} = (\bar{\eta}_0^{(\ell)}, \bar{\eta}_1^{(\ell)}) \equiv \left(\eta_0^{(\ell)}, \frac{1}{k-1} \eta_0^{(\ell)} \right). \quad (4.10)$$

Then, from Proposition 4.1, we have the following property.

Lemma 4.3 $\bar{\boldsymbol{\eta}}^{(\ell)}$ is nondecreasing in $\ell = 1, 2, \dots$

PROOF. By induction for ℓ , we will show that, for any $\ell = 1, 2, \dots$,

$$(\eta_0, \eta_1) \in \mathcal{D}^{(2)}, \quad (\eta_0, \eta_1) < (\bar{\eta}_0^{(\ell)}, \bar{\eta}_1^{(\ell)}). \quad (4.14)$$

For $\ell = 1$, we first show that $(\eta_0, \eta_1) \in \mathcal{D}^{(2)}$ for $(\eta_0, \eta_1) < (\eta_0^{(1)}, \eta_1^{(1)})$. For all $U \in \mathcal{K}$, $\varphi_{0U}(\eta_0^{(1)}, \eta_1^{(1)} \mathbf{1})$ is finite since the value of φ_{0U} does not depend on the first entry (see Remark 3.2) and $\eta_1^{(1)} \leq \bar{\eta}_1^{(0)}$. From this and (4.9), for sufficiently small $\epsilon_0, \epsilon_1 > 0$ and $(\eta_0, \eta_1) = (\eta_0^{(1)} - \epsilon_0, \eta_1^{(1)} - \epsilon_1)$, we can use (a) of Lemma 3.4, and we have $\varphi(\eta_0^{(1)} - \epsilon_0, (\eta_1^{(1)} - \epsilon_1) \mathbf{1}) < \infty$. Hence, $(\eta_0, \eta_1) \in \mathcal{D}^{(2)}$ for $(\eta_0, \eta_1) < (\eta_0^{(1)}, \eta_1^{(1)})$. We are ready to obtain (4.14) for $\ell = 1$. We recall that

$$(\bar{\eta}_0^{(1)}, \bar{\eta}_1^{(1)}) = \left(\eta_0^{(1)}, \frac{1}{k-1} \eta_0^{(1)} \right). \quad (4.15)$$

In addition, the conditions in Lemma 4.2 hold with $\eta_0 < \eta_0^{(1)}$ since $\eta_1^{(1)} > 0$. Thus, from Lemma 4.2 and (4.15), we have (4.14) for $\ell = 1$.

Suppose that (4.14) satisfies for $\ell = \ell' \geq 2$. Similarly to the arguments for $\ell = 1$, we have $(\eta_0, \eta_1) \in \mathcal{D}^{(2)}$ for $(\eta_0, \eta_1) < (\eta_0^{(\ell'+1)}, \eta_1^{(\ell'+1)})$. Thus, conditions in Lemma 4.2 hold with $\eta_0 < \eta_0^{(\ell'+1)}$ again, and we obtain (4.14) for $\ell = \ell' + 1$. \square

4.3 The last step of the proof

By Lemma 4.4 and $\bar{\eta}_1^{(\infty)} = \frac{1}{k-1} \bar{\eta}_0^{(\infty)}$, it is sufficient to prove $\bar{\eta}_0^{(\infty)} = \log \rho^{-k}$. From (4.11), we clearly have

$$\eta_0^{(\infty)} \leq \log \rho^{-k}. \quad (4.16)$$

Suppose that $\bar{\eta}_0^{(\infty)} = \alpha < \log \rho^{-k}$. Then, it is easy to see that

$$\begin{aligned} \gamma_{+U}^{(2)}(\alpha, \eta_1^{(\ell')}) &= 1, & \exists U \in \mathcal{K}, |U| = 1, \\ \gamma_{+U'}^{(2)}(\alpha, \eta_1^{(\ell')}) &< 1, & \forall U' \neq U, |U'| = 1, \end{aligned}$$

for some $\ell' \geq 1$. Thus, from (4.13), we have

$$\frac{1}{k} \alpha \geq \eta_1^{(\ell')}.$$

Moreover, from Proposition 4.1 and (4.11), for any $U \in \mathcal{K}$, there exists a small $\epsilon > 0$ such that

$$\gamma_{+U}^{(2)}(\alpha - \epsilon, \eta_1^{(\ell')}) < 1.$$

From (4.10),

$$\bar{\eta}_1^{(\ell')} = \frac{1}{k-1} (\alpha - \epsilon) > \eta_1^{(\ell')}.$$

since ϵ is sufficiently small. Thus, by Proposition 4.1, there exist $\delta > 0$ and $\eta_1 \leq \eta_1^{(\ell')}$ such that

$$\gamma_{+U}^{(2)}(\alpha + \delta, \eta_1) < 1, \quad \forall U \in \mathcal{K}.$$

This is a contradiction. From this and (4.16), we obtain $\bar{\eta}_0^{(\infty)} = \log \rho^{-k}$.

5 Concluding remarks

First of all, we note that our assumption on the tie break can be relaxed. We have assumed that arriving customers choose one of the shortest queues with equal probabilities. This assumption makes our arguments simpler, but is not essential. Namely, for each configuration of the shortest queues, we can replace it by any distribution.

In this paper, for the M/M -JSQ with k parallel queues, we obtained the exact tail asymptotics of the stationary distribution for the minimum queue length given the differences of queue length between each queue and minimum queue. It may be interesting to find c_h of (2.10) to see the joint distribution of the background state in the asymptotic formula. For this, we need to derive the left invariant vector \mathbf{x} in (3.3) because c_h is proportional to the h -th entry of \mathbf{x} . For $k = 2$, this left invariant vector is obtained in [6]. However, this computation is difficult for the case $k \geq 3$. We leave it as an open problem.

In our proof of Theorem 2.1, we obtained the subset of the convergence domain \mathcal{D} . This subset is still useful as we have seen in the proof of Corollary 2.1. However, if we completely derive the domain \mathcal{D} , we could do much better job. For example, we may have another type of tail asymptotics, e.g., joint queue length and marginal distributions. However, this would be a hard problem since our reflecting random walk is multidimensional. This challenging problem is left for future work.

Another challenging problem is to generalize the M/M -JSQ to have dedicated streams. For $k = 2$, the rough asymptotics of the minimum queue length have been completely obtained for this generalized model in [8]. We can use the present formulations by a QBD process and a reflecting random walk. However, even for $k = 2$, the answer is very complicated because $A_*(z)$ of the QBD process may not be positive recurrent. Thus, the random walk approach may be more suitable for such a generalization.

Including this generalization, various modifications of the M/M -JSQ can be described by a multidimensional reflecting random walk with skip free jumps. Thus, it is very interesting to solve the tail asymptotic problems for a general multidimensional reflecting random walk. Some related results can be found in [4, 9]. The approach in this paper may be useful to get the rough and exact asymptotics of the multidimensional reflecting random walk.

Acknowledgements

The authors would like to thank the referees for their valuable comments and suggestions. This research was supported in part by Japan Society for the Promotion of Science under grant No. 24310115.

References

- [1] BREMAUD, P. (1999). *Markov chains: Gibbs field, Monte Carlo simulation and queues*. Springer, Berlin.

- [2] FOLEY, R. and McDONALD, D. (2001). Join the shortest queue: stability and exact asymptotics. *Annals of Applied Probability*, **11** 569–607.
- [3] KINGMAN, J. F. C. (1961). Two similar queues in parallel. *The Annals of Mathematical Statistics*, **32** 1314–1323.
- [4] KOBAYASHI, M. and MIYAZAWA, M. (2011). Tail asymptotics of the stationary distribution of a two dimensional reflecting random walk with unbounded upward jumps. Preprint.
- [5] LATOUCHE, G. and RAMASWAMI, V. (1999). *Introduction to matrix analytic methods in stochastic modeling*. ASA-SIAM Series on Statistics and Applied Probability, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- [6] LI, H., MIYAZAWA, M. and ZHAO, Y. Q. (2007). Geometric decay in a QBD process with countable background states with applications to a join-the-shortest-queue model. *Stoch. Models*, **23** 413–438. URL <http://dx.doi.org/10.1080/15326340701471042>.
- [7] MEYN, S. and TWEEDIE, R. L. (2009). *Markov chains and stochastic stability*. 2nd ed. Cambridge University Press, Cambridge.
- [8] MIYAZAWA, M. (2009). Two sided DQBD process and solutions to the tail decay rate problem and their applications to the generalized join shortest queue. *Advance in Queueing Theory and Network Applications*, **1** 3–33.
- [9] MIYAZAWA, M. (2011). Rejoinder on: light tail asymptotics in multidimensional reflecting processes for queueing networks. *TOP* 1–4.
- [10] MIYAZAWA, M. and ZHAO, Y. Q. (2004). The stationary tail asymptotics in the $GI/G/1$ -type queue with countably many background states. *Adv. in Appl. Probab.*, **36** 1231–1251. URL <http://dx.doi.org/10.1239/aap/1103662965>.
- [11] NEUTS, M. F. (1981). *Matrix-geometric solutions in stochastic models: an algorithm approach*. The John Hopkins University Press, Baltimore, MD.
- [12] PUHALSKII, A. A. and VLADIMIROV, A. A. (2007). A large deviation principle for join the shortest queue. *Mathematics of Operations Research*, **32** 700–710.
- [13] SAKUMA, Y. (2010). Asymptotic behavior for marp/ph/c queue with shortest queue discipline and jockeying. *Operations Research Letters*, **38** 7–10.
- [14] SAKUMA, Y. (2011). Asymptotic behavior for marp/ph/2 queue with join the shortest queue disciplin. *Journal of the Operations Research Society of Japan*, **54** 46–64.
- [15] TAKAHASHI, Y., FUJIMOTO, K. and MAKIMOTO, N. (2001). Geometric decay of the steady-state probabilities in a quasi-birth-and-death process with a countable number of phases. *Stochastic Models*, **17** 1–24.

A Proof of Lemma 3.1

Let $S(\mathbf{h}) = \{i \in J; \mathbf{h} \in \mathcal{H}, h_i = 0\}$, that is, the set of servers having minimum queue. For its proof, we give detailed form of A_i for $i = 0, \pm 1$. For $\mathbf{h}, \mathbf{h}' \in \mathcal{H}$, we recall that $[A_i]_{\mathbf{h}, \mathbf{h}'}$ is the $(\mathbf{h}, \mathbf{h}')$ -th block of A_i .

- (a) Entries of A_{+1} : In this case, the minimum queue length increases by 1. This implies $|S(\mathbf{h})| = 1$. If $S(\mathbf{h}) = \{i\}$, arriving customers join the queue i , and the difference of queue length between queue j and the minimum queue decreases by 1 for $j \neq i$. Hence, we have

$$[A_{+1}]_{\mathbf{h}, \mathbf{h} - \mathbf{e}_i} = \lambda, \quad \mathbf{h} \in \mathcal{H}, |S(\mathbf{h})| = 1, i \in S(\mathbf{h}).$$

- (b) Entries of A_0 : The level is unchanged. When a customer arrives, it is required that $|S(\mathbf{h})| \neq 1$ and the arriving customer joins queue $i \in S(\mathbf{h})$ with probability $|S(\mathbf{h})|^{-1}$, which changes the background state from \mathbf{h} to $\mathbf{h} + \mathbf{e}_i$. Hence, we have

$$[A_0]_{\mathbf{h}, \mathbf{h} + \mathbf{e}_i} = |S(\mathbf{h})|^{-1} \lambda, \quad \mathbf{h} \in \mathcal{H}, |S(\mathbf{h})| \neq 1, i \in S(\mathbf{h}).$$

When a customer finishes service, it must be at queue $j \in J \setminus S(\mathbf{h})$, by which the difference of queue length between queue j and the shortest queue decreases by 1. Hence, we have

$$[A_0]_{\mathbf{h}, \mathbf{h} - \mathbf{e}_j} = \mu_j, \quad \mathbf{h} \in \mathcal{H}, j \in J \setminus S(\mathbf{h}).$$

- (c) Entries of A_{-1} : This case implies minimum queue length decreases by 1. That is, service completes at queue $i \in S(\mathbf{h})$, by which the background state changes from \mathbf{h} to $\mathbf{h} + \mathbf{1} - \mathbf{e}_i$. Hence, we have

$$[A_{-1}]_{\mathbf{h}, \mathbf{h} + \mathbf{1} - \mathbf{e}_i} = \mu_i, \quad \mathbf{h} \in \mathcal{H}, i \in S(\mathbf{h}).$$

PROOF OF LEMMA 3.1 For $\mathbf{h} \in \mathcal{H}$, $\mathbf{y} = (y_{\mathbf{h}})$ and $S(\mathbf{h}) \in \mathcal{K}$ satisfying $|S(\mathbf{h})| = 1$ and $i \in S(\mathbf{h})$, from (a), (b) and (c), the \mathbf{h} -th element of $A_*(z)\mathbf{y}$ is given by

$$[A_*(z)\mathbf{y}]_{\mathbf{h}} = z^{-1} \mu_i y_{\mathbf{h} + \mathbf{1} - \mathbf{e}_i} + \sum_{j \neq i} \mu_j y_{\mathbf{h} - \mathbf{e}_j} + z \lambda y_{\mathbf{h} - \mathbf{1} + \mathbf{e}_i}.$$

Substituting $z = \rho^{-k}$ and $y_{\mathbf{h}} = \rho^{-\mathbf{h}\mathbf{1}}$ into this equation, we have

$$\begin{aligned} [A_*(\rho^{-k})\mathbf{y}]_{\mathbf{h}} &= \rho^k \mu_i \rho^{-\mathbf{h}\mathbf{1} - (k-1)} + \sum_{j \neq i} \mu_j \rho^{-\mathbf{h}\mathbf{1} + 1} + \rho^{-k} \lambda \rho^{-\mathbf{h}\mathbf{1} + (k-1)} \\ &= \rho^{-\mathbf{h}\mathbf{1}} \left(\sum_{j=1}^k \mu_j \rho + \lambda \rho^{-1} \right) \\ &= y_{\mathbf{h}}, \end{aligned}$$

where we use our assumption (2.2) for the last equality.

We next consider $2 \leq |S(\mathbf{h})| \leq k$. Then, from (b) and (c), we have, for $\mathbf{h} \in \mathcal{H}$,

$$[A_*(z)\mathbf{y}]_{\mathbf{h}} = \sum_{i \in S(\mathbf{h})} z^{-1} \mu_i \mathbf{y}_{\mathbf{h}+\mathbf{1}-\mathbf{e}_i} + \sum_{j \in J \setminus S(\mathbf{h})} \mu_j \mathbf{y}_{\mathbf{h}-\mathbf{e}_j} + \sum_{i \in S(\mathbf{h})} \frac{1}{|S(\mathbf{h})|} \lambda \mathbf{y}_{\mathbf{h}+\mathbf{e}_i}.$$

Hence, we have

$$\begin{aligned} [A_*(\rho^{-k})\mathbf{y}]_{\mathbf{h}} &= \sum_{i \in S(\mathbf{h})} \rho^k \mu_i \rho^{-\mathbf{h}\mathbf{1}-(k-1)} + \sum_{j \notin S(\mathbf{h})} \mu_j \rho^{-\mathbf{h}\mathbf{1}+1} + \lambda \rho^{-\mathbf{h}\mathbf{1}-1} \\ &= \rho^{-\mathbf{h}\mathbf{1}} \left(\sum_{j=1}^k \mu_j \rho + \lambda \rho^{-1} \right) \\ &= y_{\mathbf{h}}. \end{aligned}$$

□

B Proof of Lemma 3.2

From the Foster's theorem (see e.g., [1, 7]), for some $\epsilon > 0$, it is sufficient to find a function f and finite set $\mathcal{F} \subset \mathcal{H}$ such that

$$\inf_{\mathbf{h} \in \mathcal{H}} f(\mathbf{h}) > -\infty, \quad (\text{B.1})$$

$$\sum_{\mathbf{h}' \in \mathcal{H}} p_{\mathbf{h},\mathbf{h}'} f(\mathbf{h}') < \infty \quad \text{for all } \mathbf{h} \in \mathcal{F}, \quad (\text{B.2})$$

$$\sum_{\mathbf{h}' \in \mathcal{H}} p_{\mathbf{h},\mathbf{h}'} f(\mathbf{h}') - f(\mathbf{h}) \leq -\epsilon \quad \text{for all } \mathbf{h} \in \mathcal{H} \setminus \mathcal{F}, \quad (\text{B.3})$$

where $p_{\mathbf{h},\mathbf{h}'}$ are the $(\mathbf{h}, \mathbf{h}')$ -th entry of $\Delta_{\mathbf{y}}^{-1} A_*(\rho^{-k}) \Delta_{\mathbf{y}}$. The entries of the stochastic matrix $\Delta_{\mathbf{y}}^{-1} A_*(\rho^{-k}) \Delta_{\mathbf{y}}$ is given by the following forms.

- For $|S(\mathbf{h})| = 1$,

$$p_{\mathbf{h},\mathbf{h}'} = \begin{cases} \lambda \rho^{-1}, & \mathbf{h}' = \mathbf{h} - \mathbf{1} + \mathbf{e}_i, i \in S(\mathbf{h}), \\ \mu_i \rho, & \mathbf{h}' = \mathbf{h} + \mathbf{1} - \mathbf{e}_i, i \in S(\mathbf{h}) \\ & \text{or } \mathbf{h}' = \mathbf{h} - \mathbf{e}_i, i \in J \setminus S(\mathbf{h}), \\ 0, & \text{otherwise.} \end{cases}$$

- For $2 \leq |S(\mathbf{h})| \leq k$,

$$p_{\mathbf{h},\mathbf{h}'} = \begin{cases} \frac{\lambda \rho^{-1}}{|S(\mathbf{h})|}, & \mathbf{h}' = \mathbf{h} + \mathbf{e}_i, i \in S(\mathbf{h}), \\ \mu_i \rho, & \mathbf{h}' = \mathbf{h} + \mathbf{1} - \mathbf{e}_i, i \in S(\mathbf{h}) \\ & \text{or } \mathbf{h}' = \mathbf{h} - \mathbf{e}_i, i \in J \setminus S(\mathbf{h}), \\ 0, & \text{otherwise.} \end{cases}$$

For $\mathbf{h} = (h_1, h_2, \dots, h_k) \in \mathcal{H}$, let

$$f(\mathbf{h}) = \frac{1}{2} \sum_{j=1}^{k-1} \sum_{m=j+1}^k (h_j - h_m)^2.$$

For this f , we obviously have (B.1). We note that $i \in S(\mathbf{h})$ implies $h_i = 0$. For $i \in S(\mathbf{h})$ and $\mathbf{h}' = \mathbf{h} - \mathbf{1} + \mathbf{e}_i$ or $\mathbf{h}' = \mathbf{h} + \mathbf{e}_i$, we have

$$(h'_j - h'_m)^2 = \begin{cases} (h_j - h_m)^2, & j, m \neq i, \\ (h_j - h_m)^2 - (2h_j - 1), & j < m = i, \\ (h_j - h_m)^2 - (2h_m - 1), & j = i < m. \end{cases}$$

Thus, we have

$$\begin{aligned} p_{\mathbf{h}, \mathbf{h}'} f(\mathbf{h}') &= \frac{\lambda \rho^{-1}}{2|S(\mathbf{h})|} \left(\sum_{j=1}^{k-1} \sum_{m=j+1}^k (h_j - h_m)^2 - \sum_{j=1}^{i-1} (2h_j - 1) - \sum_{m=i+1}^k (2h_m - 1) \right) \\ &= \frac{\lambda \rho^{-1}}{|S(\mathbf{h})|} \left(f(\mathbf{h}) - \sum_{j=1}^k h_j + \frac{k-1}{2} \right). \end{aligned}$$

Similarly, for $\mathbf{h}' = \mathbf{h} + \mathbf{1} - \mathbf{e}_i$ and $i \in S(\mathbf{h})$,

$$p_{\mathbf{h}, \mathbf{h}'} f(\mathbf{h}') = \mu_i \rho \left(f(\mathbf{h}) + \sum_{j=1}^k h_j + \frac{k-1}{2} \right).$$

For $\mathbf{h}' = \mathbf{h} - \mathbf{e}_i$ and $i \in J \setminus S(\mathbf{h})$, we also obtain,

$$(h'_j - h'_m)^2 = \begin{cases} (h_j - h_m)^2, & j, m \neq i, \\ (h_j - h_m)^2 + 2(h_j - h_m) + 1, & j < m = i, \\ (h_j - h_m)^2 + 2(h_m - h_j) + 1, & j = i < m, \end{cases}$$

so we have the following inequality.

$$\begin{aligned} p_{\mathbf{h}, \mathbf{h}'} f(\mathbf{h}') &= \mu_i \rho \left(f(\mathbf{h}) + \sum_{j=1}^{i-1} \left(h_j - h_i + \frac{1}{2} \right) + \sum_{m=i+1}^k \left(h_m - h_i + \frac{1}{2} \right) \right) \\ &\leq \mu_i \rho \left(f(\mathbf{h}) + \sum_{j=1}^k h_j + \frac{k-1}{2} \right). \end{aligned}$$

From (2.2), we have

$$\begin{aligned}
\sum_{\mathbf{h}' \in \mathcal{H}} p_{\mathbf{h}, \mathbf{h}'} f(\mathbf{h}') &\leq \sum_{i \in S(\mathbf{h})} \frac{\lambda \rho^{-1}}{|S(\mathbf{h})|} \left(f(\mathbf{h}) - \sum_{j=1}^k h_j + \frac{k-1}{2} \right) \\
&\quad + \sum_{i \in S(\mathbf{h})} \mu_i \rho \left(f(\mathbf{h}) + \sum_{j=1}^k h_j + \frac{k-1}{2} \right) \\
&\quad + \sum_{i \in J \setminus S(\mathbf{h})} \mu_i \rho \left(f(\mathbf{h}) + \sum_{j=1}^k h_j + \frac{k-1}{2} \right) \\
&= f(\mathbf{h}) + \left(\lambda - \sum_{j=1}^k \mu_j \right) \mathbf{h} \mathbf{1} + \frac{k-1}{2}.
\end{aligned}$$

Thus, from the stability condition (2.1), (B.2) and (B.3) hold with

$$\mathcal{F} = \left\{ \mathbf{h} \in \mathcal{H}; \mathbf{h} \mathbf{1} \leq \frac{k-1+\epsilon}{2(\sum_{i=1}^k \mu_i - \lambda)} \right\}.$$

We complete the proof.

C Proof of Lemma 3.3

It suffices to prove that there exists α such that

$$\mathbb{E}(e^{\alpha \sqrt{\sum_{i=1}^k L_i^2}}) < \infty, \quad (\text{C.1})$$

since $\mathbb{E}(e^{\frac{\alpha}{\sqrt{k}} \sum_{i=1}^k L_i}) \leq \mathbb{E}(e^{\alpha \sqrt{\sum_{i=1}^k L_i^2}})$. To prove (C.1), we consider a Lyapunov function such that

$$f(\mathbf{u}) = e^{\alpha \sqrt{\sum_{i=1}^k u_i^2}},$$

for $\mathbf{u} = (u_1, u_2, \dots, u_k) \in \mathbb{Z}_+^k$, and, for each $\beta > 0$, define

$$F_\beta = \left\{ \mathbf{u} \in \mathbb{Z}_+^k; \sum_{i=1}^k u_i \leq \beta \right\}.$$

Obviously, F_β is a finite set. To verify (C.1), if we can show that

$$\mathbb{E}(f(\mathbf{L}_{n+1}) | \mathbf{L}_n = \mathbf{u}) - f(\mathbf{u}) \leq -c_1 f(\mathbf{u}), \quad \forall \mathbf{u} \in \mathbb{Z}_+^k \setminus F_\beta, \quad (\text{C.2})$$

for some $c_1 > 0$, then, for any $\mathbf{u} \in \mathbb{Z}_+^k$,

$$\mathbb{E}(f(\mathbf{L}_{n+1}) | \mathbf{L}_n = \mathbf{u}) - (1 - c_1) f(\mathbf{u}) \leq s(\mathbf{u}),$$

where s is a function defined

$$s(\mathbf{u}) = (\mathbb{E}(f(\mathbf{L}_{n+1}) | \mathbf{L}_n = \mathbf{u}) - (1 - c_1) f(\mathbf{u}))^+ \mathbf{1}(\mathbf{u} \in F_\beta).$$

Thus, by Theorem 14.3.7 of [7], we have

$$c_1 \sum_{\mathbf{u} \in \mathbb{Z}_+^k} f(\mathbf{u}) \mathbb{P}(\mathbf{L} = \mathbf{u}) = c_1 \mathbb{E}(e^{\alpha \sqrt{\sum_{i=1}^k L_i^2}}) \leq \sum_{\mathbf{u} \in \mathbb{Z}_+^k} s(\mathbf{u}) \mathbb{P}(\mathbf{L} = \mathbf{u}) < \infty,$$

which implies (C.1). We will show that, for $0 < \alpha < \sqrt{k} \log \rho^{-1}$, we can find $\beta > 0$ such that (C.2) holds.

For $\mathbf{u} \in \mathbb{Z}_+$, let $\mathbf{u}_\ell = (u_{\ell_1}, u_{\ell_2}, \dots, u_{\ell_k})$ be a permutation of \mathbf{u} satisfying

$$u_{\ell_1} \leq u_{\ell_2} \leq \dots \leq u_{\ell_k}. \quad (\text{C.3})$$

We exclude the case $u_{\ell_1} = u_{\ell_2} = \dots = u_{\ell_k} = 0$, that is, we assume the following condition.

$$0 < u_{\ell_j}, \quad \exists j \in J. \quad (\text{C.4})$$

Then, for $j \in J$ satisfying (C.4), it follow from the transition matrix of the original queue length process $\{\mathbf{L}_\ell; \ell \in \mathbb{Z}_+\}$ that

$$\begin{aligned} \mathbb{E}(f(\mathbf{L}_{n+1}) | \mathbf{L}_n = \mathbf{u}) - f(\mathbf{u}) &= \lambda e^{\alpha \sqrt{(u_{\ell_1}+1)^2 + \sum_{i=2}^k u_{\ell_i}^2}} + \sum_{i=j}^k \mu_{\ell_i} e^{\alpha \sqrt{(u_{\ell_i}-1)^2 + \sum_{m \neq i} u_{\ell_m}^2}} \\ &\quad + \sum_{i=1}^{j-1} \mu_{\ell_i} e^{\alpha \sqrt{\sum_{m=1}^k u_{\ell_m}^2}} - e^{\alpha \sqrt{\sum_{i=1}^k u_{\ell_i}^2}} \\ &= \lambda e^{\alpha \sqrt{(u_{\ell_1}+1)^2 + \sum_{i=2}^k u_{\ell_i}^2}} + \sum_{i=j}^k \mu_{\ell_i} e^{\alpha \sqrt{(u_{\ell_i}-1)^2 + \sum_{m \neq j} u_{\ell_m}^2}} \\ &\quad - \left(\lambda + \sum_{i=j}^k \mu_{\ell_i} \right) e^{\alpha \sqrt{\sum_{i=1}^k u_{\ell_i}^2}}, \end{aligned}$$

where we have used (2.2) to get the second equality. Note that $\sqrt{1+u} \leq 1 + \frac{u}{2}$ for $u \geq -1$, so we have the following inequalities.

$$\begin{aligned} \sqrt{(u_{\ell_1}+1)^2 + \sum_{i=2}^k u_{\ell_i}^2} &= r \sqrt{1 + \frac{1+2u_{\ell_1}}{r^2}} \\ &\leq r + \frac{1+2u_{\ell_1}}{2r}, \\ \sqrt{(u_{\ell_i}-1)^2 + \sum_{m \neq i} u_{\ell_m}^2} &= r \sqrt{1 + \frac{1-2u_{\ell_i}}{r^2}} \\ &\leq r + \frac{1-2u_{\ell_i}}{2r}, \end{aligned}$$

where $r = \sqrt{\sum_{i=1}^k u_{\ell_i}^2}$. Hence, we have

$$\begin{aligned}
& \mathbb{E}(f(\mathbf{L}_{n+1}) | \mathbf{L}_n = \mathbf{u}) - f(\mathbf{u}) \\
& \leq \lambda e^{\alpha r + \frac{\alpha(1+2u_{\ell_1})}{2r}} + \sum_{i=j}^k \mu_{\ell_i} e^{\alpha r + \frac{\alpha(1-2u_{\ell_i})}{2r}} - \left(\lambda + \sum_{i=j}^k \mu_{\ell_i} \right) e^{\alpha r} \\
& = e^{\alpha r} \left(e^{\frac{\alpha}{2r}} \left(\lambda e^{\frac{\alpha u_{\ell_1}}{r}} + \sum_{i=j}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_i}}{r}} \right) - \left(\lambda + \sum_{i=j}^k \mu_{\ell_i} \right) \right) \\
& = f(\mathbf{u}) \left(e^{\frac{\alpha}{2r}} \left(\lambda e^{\frac{\alpha u_{\ell_1}}{r}} + \sum_{i=j}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_i}}{r}} \right) - \left(\lambda + \sum_{i=j}^k \mu_{\ell_i} \right) \right).
\end{aligned}$$

From this inequality, for a small $\alpha > 0$ and a large $\beta > 0$, it is enough to find $c_1 > 0$ such that

$$e^{\frac{\alpha}{2r}} \left(\lambda e^{\frac{\alpha u_{\ell_1}}{r}} + \sum_{i=j}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_i}}{r}} \right) - \left(\lambda + \sum_{i=j}^k \mu_{\ell_i} \right) \leq -c_1, \quad \forall \mathbf{u} \in \mathbb{Z}_+ \setminus \mathcal{F}_\beta. \quad (\text{C.5})$$

For this, for a small $\delta > 0$, we partition two cases $\frac{\alpha u_{\ell_1}}{r} < \delta$ and $\frac{\alpha u_{\ell_1}}{r} \geq \delta$. From our assumption (C.3) and (C.4), we have

$$0 \leq \frac{u_{\ell_i}}{r} \leq 1, \quad i = 1, 2, \dots, k, \quad (\text{C.6})$$

$$0 \leq \frac{u_{\ell_1}}{r} \leq \frac{1}{\sqrt{k}}, \quad (\text{C.7})$$

$$\frac{1}{\sqrt{k}} \leq \frac{u_{\ell_k}}{r} \leq 1. \quad (\text{C.8})$$

First assume that $\frac{\alpha u_1}{r} < \delta$. In this case, from (C.6) and (C.8), we obtain

$$\begin{aligned}
& e^{\frac{\alpha}{2r}} \left(\lambda e^{\frac{\alpha u_{\ell_1}}{r}} + \sum_{i=j}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_i}}{r}} \right) - \left(\lambda + \sum_{i=j}^k \mu_{\ell_i} \right) \\
& \leq e^{\frac{\alpha}{2r}} \left(\lambda e^{\frac{\alpha u_{\ell_1}}{r}} + \sum_{i=j}^{k-1} \mu_{\ell_i} + \mu_{\ell_k} e^{-\frac{\alpha}{\sqrt{k}}} \right) - \left(\lambda + \sum_{i=j}^k \mu_{\ell_i} \right) \\
& = \lambda (e^{\frac{\alpha}{2r} + \delta} - 1) - \mu_{\ell_k} (1 - e^{-\frac{\alpha}{\sqrt{k}}}).
\end{aligned}$$

Then, for small α satisfying $0 < \alpha < \sqrt{k} \log \rho^{-1}$, there exist sufficiently small δ and large β such that

$$d_1 \equiv \lambda (e^{\frac{\alpha}{2\sqrt{\beta}} + \delta} - 1) - \mu_{\ell_k} (1 - e^{-\frac{\alpha}{\sqrt{k}}}) < 0. \quad (\text{C.9})$$

We next consider the case $\frac{\alpha u_{\ell_1}}{r} \geq \delta$. Then, it implies that $u_1 > 0$. From (C.3) and

(C.7),

$$\begin{aligned}
& e^{\frac{\alpha}{2r}} \left(\lambda e^{\frac{\alpha u_{\ell_1}}{r}} + \sum_{i=j}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_i}}{r}} \right) - \left(\lambda + \sum_{i=j}^k \mu_{\ell_i} \right) \\
&= (1 + e^{\frac{\alpha}{2r}} - 1) \left(\lambda e^{\frac{\alpha u_{\ell_1}}{r}} + \sum_{i=1}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_i}}{r}} \right) - \left(\lambda + \sum_{i=1}^k \mu_{\ell_i} \right) \\
&\leq (1 + e^{\frac{\alpha}{2r}} - 1) \left(\lambda e^{\frac{\alpha u_{\ell_1}}{r}} + \sum_{i=1}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_1}}{r}} \right) - \left(\lambda + \sum_{i=1}^k \mu_{\ell_i} \right) \\
&= \left(e^{\frac{\alpha u_{\ell_1}}{r}} - 1 \right) \left(\lambda - \sum_{i=1}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_1}}{r}} \right) + (e^{\frac{\alpha}{2r}} - 1) \left(\lambda e^{\frac{\alpha u_{\ell_1}}{r}} + \sum_{i=1}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_1}}{r}} \right) \\
&\leq \left(e^{\frac{\alpha u_{\ell_1}}{r}} - 1 \right) \left(\lambda - \sum_{i=1}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_1}}{r}} \right) + (e^{\frac{\alpha}{2r}} - 1) \left(\lambda e^{\frac{\alpha}{\sqrt{k}}} + \sum_{i=1}^k \mu_{\ell_i} \right) \\
&\leq (e^{\delta} - 1) \left(\lambda - \sum_{i=1}^k \mu_{\ell_i} e^{-\frac{\alpha}{\sqrt{k}}} \right) + (e^{\frac{\alpha}{2r}} - 1) \left(\lambda e^{\frac{\alpha}{\sqrt{k}}} + \sum_{i=1}^k \mu_{\ell_i} \right),
\end{aligned}$$

where the last inequality is given by $\frac{\alpha u_1}{r} \geq \delta$ and

$$\lambda - \sum_{i=1}^k \mu_{\ell_i} e^{-\frac{\alpha u_{\ell_1}}{r}} \leq \lambda - \sum_{i=1}^k \mu_{\ell_i} e^{-\frac{\alpha}{\sqrt{k}}} < 0$$

for $0 < \alpha < \sqrt{k} \log \rho^{-1}$. Thus, for fixed α and δ , we obtain (C.5) for sufficiently large $\beta > 0$ satisfying

$$d_2 \equiv \left(e^{\frac{\alpha}{2\sqrt{\beta}}} - 1 \right) \left(\lambda e^{\frac{\alpha}{\sqrt{k}}} - \sum_{i=1}^k \mu_{\ell_i} \right) + (e^{\delta} - 1) \left(\lambda - \sum_{i=1}^k \mu_{\ell_i} e^{-\frac{\alpha}{\sqrt{k}}} \right) < 0. \quad (\text{C.10})$$

We put $c_1 = -\max(d_1, d_2)$. Then, from (C.9) and (C.10), we have $c_1 > 0$ and (C.5). This completes the proof since $\varphi(0, \theta \mathbf{1}) \leq \mathbb{E}(e^{\theta \sum_{i=1}^k L_i})$ for any $\theta \geq 0$.

D Proof of Lemma 3.4

We prove (3.16). For this, we apply truncation argument for the moment generating functions φ_{0U} and φ_{+U} . For each $n = 1, 2, \dots$, let

$$g_n(x) = \min(x, n), \quad x \in \mathbb{R}.$$

Then, for any $x, y \in \mathbb{R}$, it is easy to see that

$$g_n(x + y) \leq g_n(x) + \begin{cases} y & x \leq n, \\ 0 & x > n. \end{cases} \quad (\text{D.1})$$

The moment generating function $E(e^{g_n(\boldsymbol{\theta}\mathbf{Z})})$ is finite for each n and any $\boldsymbol{\theta} \in \mathbb{R}^{k+1}$. So, we have, by the stationary equation (2.9),

$$\mathbb{E}(e^{g_n(\boldsymbol{\theta}\mathbf{Z})}) = \sum_{U \in \mathcal{K}} \left(\mathbb{E} \left(e^{g_n(\boldsymbol{\theta}\mathbf{Z} + \boldsymbol{\theta}\mathbf{X}^{(0U)})} 1(\mathbf{Z} \in \mathcal{S}_{0U}) \right) + \mathbb{E} \left(e^{g_n(\boldsymbol{\theta}\mathbf{Z} + \boldsymbol{\theta}\mathbf{X}^{(+U)})} 1(\mathbf{Z} \in \mathcal{S}_{+U}) \right) \right).$$

Since $\mathbf{X}^{(+U)}$ is independent for \mathbf{Z} , from (D.1), we have, for any $U \in \mathcal{K}$,

$$\begin{aligned} \mathbb{E} \left(e^{g_n(\boldsymbol{\theta}\mathbf{Z} + \boldsymbol{\theta}\mathbf{X}^{(+U)})} 1(\mathbf{Z} \in \mathcal{S}_{+U}) \right) &\leq \mathbb{E} \left(e^{g_n(\boldsymbol{\theta}\mathbf{Z}) + \boldsymbol{\theta}\mathbf{X}^{(+U)}} 1(\mathbf{Z} \in \mathcal{S}_{+U}, \boldsymbol{\theta}\mathbf{Z} \leq n) \right) \\ &\quad + \mathbb{E} \left(e^{g_n(\boldsymbol{\theta}\mathbf{Z})} 1(\mathbf{Z} \in \mathcal{S}_{+U}, \boldsymbol{\theta}\mathbf{Z} > n) \right) \\ &= \gamma_{+U}(\boldsymbol{\theta}) \mathbb{E} \left(e^{g_n(\boldsymbol{\theta}\mathbf{Z})} 1(\mathbf{Z} \in \mathcal{S}_{+U}, \boldsymbol{\theta}\mathbf{Z} \leq n) \right) \\ &\quad + \mathbb{E} \left(e^{g_n(\boldsymbol{\theta}\mathbf{Z})} 1(\mathbf{Z} \in \mathcal{S}_{+U}, \boldsymbol{\theta}\mathbf{Z} > n) \right). \end{aligned}$$

We have similar result for $\mathbf{X}^{(0U)}$. By the decomposition of $\mathbb{E}(e^{g_n(\boldsymbol{\theta}\mathbf{Z})})$,

$$\begin{aligned} &\sum_{U \in \mathcal{K}} (1 - \gamma_{+U}(\boldsymbol{\theta})) \mathbb{E} \left(e^{g_n(\boldsymbol{\theta}\mathbf{Z})} 1(\mathbf{Z} \in \mathcal{S}_{+U}, \boldsymbol{\theta}\mathbf{Z} \leq n) \right) \\ &\leq \sum_{U \in \mathcal{K}} (\gamma_{0U}(\boldsymbol{\theta}) - 1) \mathbb{E} \left(e^{g_n(\boldsymbol{\theta}\mathbf{Z})} 1(\mathbf{Z} \in \mathcal{S}_{0U}, \boldsymbol{\theta}\mathbf{Z} \leq n) \right). \end{aligned}$$

Hence, we obtain (3.16) as $n \rightarrow \infty$ by the monotone convergence theorem. We complete the proof since we can use a similar argument to (3.19).

E Proof of Corollary 2.1

As we already noted, we only need to prove that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(M = n) \leq \log \rho^k. \quad (\text{E.1})$$

From Lemma 3.5, we have, for $\eta_0 < \log \rho^{-k}$ and $0 \leq \eta_1 < \frac{1}{k-1} \log \rho^{-k}$,

$$e^{\eta_0 n} \mathbb{P}(M = n) \leq \mathbb{E}(e^{\eta_0 M}) \leq \varphi(\eta_0, \eta_1 \mathbf{1}) < \infty.$$

Hence,

$$\eta_0 n + \log \mathbb{P}(M = n) \leq \log \varphi(\eta_0, \eta_1 \mathbf{1}),$$

which implies that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(M = n) \leq -\eta_0.$$

Thus, letting $\eta_0 \uparrow \log \rho^{-k}$ yields (E.1).